

Estimación de la integridad ecosistémica de las costas arenosas mexicanas a través de técnicas de aprendizaje de máquina

Proyecto Conahcyt de ciencia de frontera (CF-2023-G-1497)

Octavio Pérez-Maqueo Miguel Equihua

2024-05-23

Tabla de contenidos

Preface

Este es el documento de registro de avances del proyecto **Estimación de la integridad ecosistémica de las costas arenosas mexicanas a través de técnicas de aprendizaje de máquina**

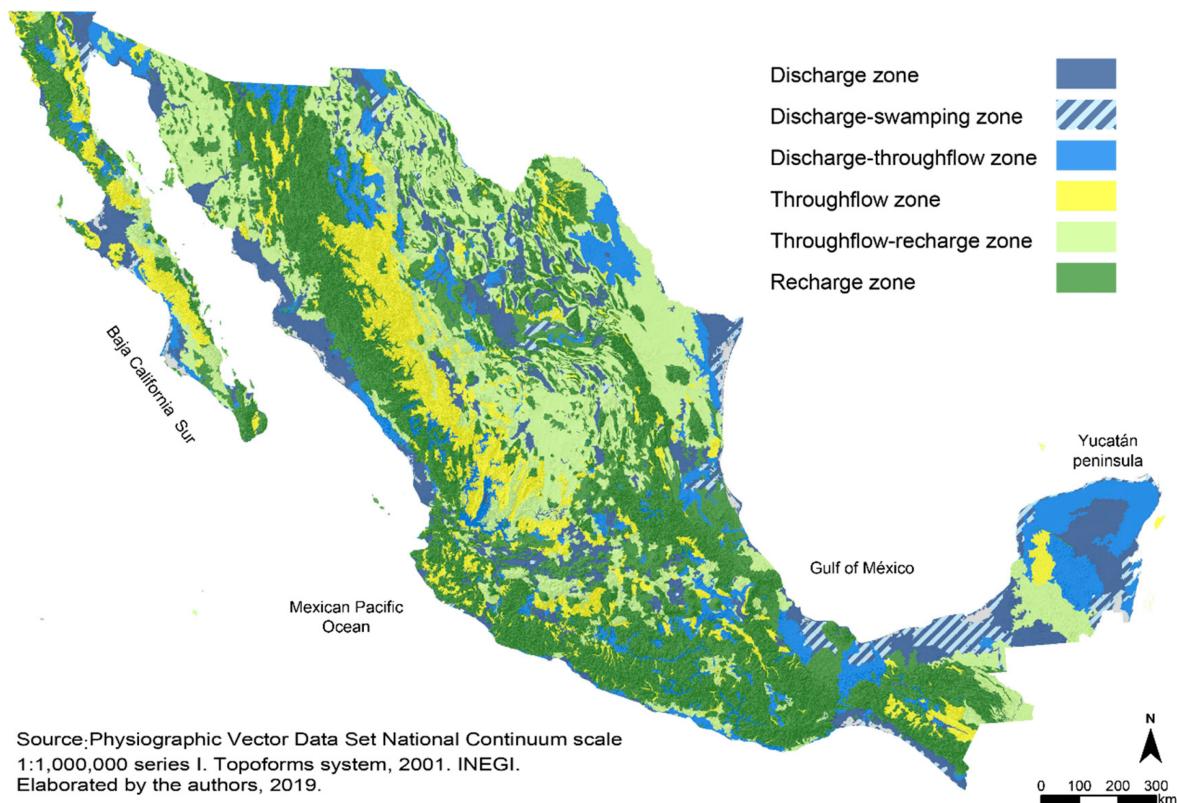
1 Resumen

México, como otros países, inició la implementación de un sistema de cuentas de ecosistemas cuyo propósito es calcular y poner regularmente a disposición de la sociedad información sobre la extensión y la condición de los ecosistemas, así como de la capacidad para obtener los diversos beneficios que la naturaleza provee como servicios ecosistémicos. Como parte de un piloto mundial, México desarrolló un sistema de contabilidad para ecosistemas terrestres (INEGI 2021). Ahora, hay el interés en extender el sistema para incluir los ambientes costeros, marinos y acuáticos continentales. Para los ecosistemas terrestres México desarrolló una propuesta que aprovecha datos masivos (Big data) y la aplicación de técnicas de inteligencia artificial (IA). La experiencia mostró que este planteamiento puede ser de enorme ayuda para llevar a cabo una contabilidad de ecosistemas a gran escala. Las técnicas de IA son atípicas dentro del ámbito de la contabilidad ambiental y por tanto se estima que la estrategia iniciada en nuestro país es innovadora. Además, las nuevas propuestas en el uso de las técnicas de aprendizaje de máquina apuntan a preferir formulaciones causales que si bien son procesadas por máquinas, produzcan resultados entendibles para los usuarios humanos. Esto es lo que está emergiendo como “Inteligencia artificial interpretable” (IAI por las siglas en inglés, Mihaljevic et al 2021). Una aproximación del tipo IAI para el caso de las cuentas de ecosistemas permitiría, con base en datos y un enfoque sistémico, un mayor entendimiento de las relaciones entre las acciones de origen antrópico y su repercusión en los ecosistemas del país y en los servicios que estos proporcionan. En este proyecto, se propone evaluar como prueba de concepto, la condición de las costas arenosas (playas y dunas) de todo el país, por medio del cálculo de un índice de integridad de los ecosistemas de costas arenosas (IIECA) bajo el paradigma de IAI.

2 Introducción

El proyecto **Estimación de la integridad ecosistémica de las costas arenosas mexicanas a través de técnicas de aprendizaje de máquina** se propone generar estimaciones de la condición en la que se encuentran los ecosistemas costeros, especialmente los arenosos, del país con una representación raster y resolución geográfica de pixeles de 3 segundos (aproximadamente 100 m). También se está considerando tener una valoración de mayor resolución para la zona costera de Quintana Roo, en donde se propone lograr una resolución geográfica con pixeles 0.3 segundos.

3 Datos



Characterization of Surface Evidence of Groundwater Flow Systems

3.1 Agenda

La reunión de inicio del proyecto (**CF-2023-G-1497**):

Estimación de la integridad ecosistémica de las costas arenosas mexicanas a través de técnicas de aprendizaje de máquina

Se realizó del miércoles 29 al viernes 31 de mayo de 2024.

3.1.1 Día 1 miércoles 29 de mayo

- **10:00-10:20 inicio**
 - bienvenida
 - presentación del taller
 - presentación de asistentes
- **10:20-11:20 ¿por qué creemos que es necesario un proyecto como este?**
 - visión integral
 - caso particular de las costas
 - desarrollo de habilidades
 - ¿cómo sugerimos hacerlo?
 - co-construcción
 - ciencia abierta
 - principios fair
 - análisis con base en propuesta del seea ea
 - flujos de trabajo (pipelines)
- **11:20-12:20 descripción y avances del proyecto**
 - generalidades
 - marco conceptual
 - compromisos para las tres etapas
 - avances
- **12:20-12:40 descanso**
- **12:40-14:00 el sistema de contabilidad de ecosistemas.**
 - activos ecosistémicos
 - cuentas de extensión
 - condición
 - servicios ecosistémicos
 - cuenta de condición
 - modelo de integridad ecosistémica
- **14:00-15:00 comida**
- **15:00-17:00 estrategia metodológica**
 - de lo conceptual a lo analítico (de miro a nética)
 - enfoque causal
 - probabilístico
 - documentación
 - gestión de datos

3.1.2 Día 2 jueves 30 de mayo

- 10:00 -14:00 gestión de datos y modelación participativa
 - variables existentes
 - nuevas variables
 - sugerencias para documentar
 - gestionar datos.
 - ejercicio de modelación participativa
- 14:00-15:00 comida y foto grupal
- 15:00-17:00 acuerdos y siguientes pasos

3.1.3 Día 3 viernes 31 de mayo

- 10:00 -12:20 Evaluación integral en las costas de Quintana Roo
- 12:20 -14:00 Acuerdos y siguientes pasos
- 14:00 Fin del taller

3.2 Miro

3.2.1 Construcción colaborativa de Redes Bayesianas

La propuesta que hacemos es usar *Miro* como plataforma de debate y construcción de consenso. Buscamos así aproximarnos a la estructura causal que vincula las variables que determinan el rendimiento y la sustentabilidad (social, económica y ambiental). Una vez resuelta esta etapa, es necesario transferir lo acordado en *Miro* a plataformas de análisis estadístico. Aquí, nos propusimos trasladar los resultados de *Miro* a **R**. Al hacer esto las posibilidades analíticas se potencian muy ampliamente. Como veremos, podemos recurrir a *dagitty* o incluso utilizar *Python* y *Netica*.



Hemos preparado una biblioteca con las rutinas necesarias para transferir lo trabajado en **Miro** (básicamente desarrollo conceptual) a **R** en donde se podrá continuar con el desarrollo formal

del modelo según se requiera, cualitativa y cuantitativamente. Para instalarla y usarla hay que seguir las [indicaciones que se describen aquí](#). Consideren que para usarla habrá que hacer algunas preparaciones para interactuar con **Miro**. Se trata básicamente de tener los permisos de acceso. Habrá que obtener credenciales en la forma de un *token*, el cual hay que evitar distribuir. Por lo tanto, para operar hemos recurrido a la biblioteca `keyring` para mantenerlas confidenciales. Esta biblioteca permite utilizar el sistema de seguridad del equipo para la gestión de las claves. Para dar de alta una llave utiliza la instrucción de registro `key_set`, necesitarás dos identificadores de tu preferencia que deberán ir entrecomillados:

```
key_set(service = "[ref. de tu preferencia]", username = "[id. de tu gusto]")
```

Una vez registrada en el equipo la llave, queda lista para su uso en cualquier momento en esa máquina y deberás proporcionar los dos identificadores que elegiste para operar las *ingesta* a **R**, del modelo hecho en **Miro**.

```
if (!require(miro2bayes))
{
  library(devtools)
  install_github("equihuam/miro2bayesNet", force = TRUE)
}

library(miro2bayes)
library(bnlearn)
library(bnviewer)
library(tidyverse, quietly = TRUE, warn.conflicts = FALSE)

tableros <- miroBoards(servMiro = "miro", user = "miguel-edu-token")
tableros[, c("name", "id")]
```

	name	id
1	Costa arenosa	uXjVKB6PRSY=
2	Education Plan Basics	uXjVPZCsPDE=
3	Education Plan Apps & Integrations	uXjVPZCsPUo=
4	Education Plan Administration	uXjVPZCsPVc=
5	Education Plan Useful Resources	uXjVPZCsPVQ=

```
tablero_tr <- tableros %>%
  filter(str_detect(name, "Costa")) %>%
  select(id, name)

datos_miro <- getMiro(servMiro = "miro", user = "miguel-edu-token",
                      board = tablero_tr)
```

```
miroValidation(datos_miro)
```

```
Miro board origin:      Costa arenosa
Is it a TRUE DAG?:    Graph is acyclic
Number of sticky notes: 46
Nodes without var:     4
Number of linked nodes: 15
Duplicated nodes:      0
Number of arcs:        13
Well connected arcs:   13
Numb. Loose arcs:      0
Duplicated arcs:       0
```

```
#neticaMiro <- miro2DNE(datos_miro)

#write(neticaMiro, "costa-arenosa.dne")

netMiro_bn <- miro2bnlearn(datos_miro)
netMiro_bn
```

Random/Generated Bayesian network

```
model:
[redondez] [vel_viento] [dir_viento] [pot_ola] [dir_ola] [d50] [lluvia_anual]
[secas] [mareas] [dir_costa] [radiacion] [flujo_subterraneo] [corrientes_m]
[erosibilidad|redondez:vel_viento:dir_viento:pot_ola:dir_ola:d50:lluvia_anual:secas:mareas]
[aguas_disp|flujo_subterraneo]

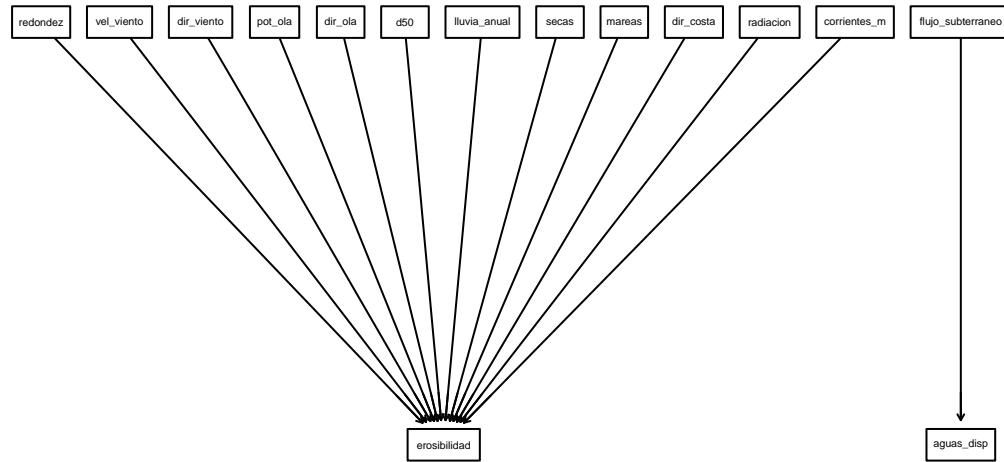
nodes:                      15
arcs:                       13
  undirected arcs:          0
  directed arcs:            13
average markov blanket size: 10.53
average neighbourhood size:  1.73
average branching factor:   0.87

generation algorithm:           Empty
```

```
variables <- tibble(var = datos_miro$nodes$var)
```

```
graphviz.plot(netMiro_bn, layout = "dot")
```

Loading required namespace: Rgraphviz



La lista de variables identificadas es:

text	var	color
potencia de la		
ola (datos	pot_ola	blue
horarios 1940-		
2022) https://confluence.ecmwf.int/display/CKB/ERA5%3A&#43;data&#43;documentation		
mareasMareografos		
de Pacífico y		
Atlántico	mareas	blue
(SEMAR)		
RODOLFO		

text	var	color
forma de		
crecimiento de		
vegetación de		
dunas:Tomar		
de base de		
datos de		
lleana/conabio.	habito	dark_green
Zonas con		
arbustivas		
tienden mayor		
erosión. en		
zonas como		
dice marisa		
eso no pasa		
disponinibilidad		
de arena	ancho_playa	dark_green
-ancho playa		
geología(geomorfología, tipo		
de roca)Pedir		
carta geológica		
minera servicio		
geologico		
mexicano.		
Octavio y	geologia	blue
Karla en		
formato shape		
"Geológica		
minera" https://www.sgm.gob.mx/CartasDisponibles/		
(en pdf hasta		
1:50000) https://www.sgm.gob.mx/GeoInfoMexGobMx/#(shape)		
Arrecifes de		
coral-Buscar la		
mejor capa:	coral	dark_green
blaqueamiento,		
Copernicus o		
Conabio https://simar.conabio.gob.mx/explorer/?undefined		
pastos	pastos	dark_green
marinos https://simar.conabio.gob.mx/explorer/?undefined		

text	var	color
Disponibilidad de agua salobre o dulce	aguas_disp	blue
Flujo subterráneo.		
Información disponible en SHP (polígonos).		
Solicitar a Rodolfo / Valeria. Tiene una clasificación. Nivel nacional.	flujo_subterrane	blue
Artículo: https://www.mdpi.com/2073-4441/12/9/2459		
Sol. WB_Global solar atlas. https://globalsolarradiation.info/map?lat=19.248922,-97.171326&lon=19.519781,-96.911168&m=site		
EvapotranspiraciónEn terminos de germinación y cobertura	evapotranspiraci	dark_green
NOTAS GENERALESJuntaremos playa y duna como un solo ecosistema. Sugerencia de Rodolfo	notas_1	gray

text	var	color
Plataforma continental. GEBCO.		
Gridded Bathymetry data download.https://download.gebco.net	plataforma_contblue	
Infraestructura	infrastruct	orange
Manglares Moreno	SIMARPaty mangle	dark_green
Estructuras marinas	estruct_marina	orange
Influencia antrópica	influ_antro	orange
ALTERACIONES		
ANTROPICAS DE LA Calidad del agua	alt_antro	orange
Pendiente de la playa puntualMónica y otros	información pendiente	dark_green
Diámetro mediano D50(base de datos GICP-IIUNAM)	d50	blue
Redondez sedimento. (base de datos GICP-IIUNAM)	redondez	blue

text	var	color
Dirección de la línea de costa con relación al oleaje y al vientoUsar	dir_costa	blue
línea de costa propuesta por Rodolfo		
Efecto del sargazo	EfectoDelSargazo	blue
Corrientes MarinasAtlas de corrientes IMTA de Efraín	corrientes_m	blue
Mateos (Base de datos del CEMIE)VALERIA		
Especies invasorasUsar datos	invasoras	dark_green
procesados por Julián		

text	var	color
<p>Perfil de sueloPerfiles edafológicos se pueden encontrar en:https://www.inegi.org.mx/app/biblioteca/ficha.html?upc&#61;702825266707Suelos sueltos y poco compactados son más susceptibles a la erosión eólica, lo que facilita la formación de dunas. Por otro lado, suelos más compactos pueden resistir mejor la acción perfilSuelo dark_green del viento. La presencia de diferentes horizontes en el perfil edafológico puede afectar la disponibilidad de sedimentos para la formación de dunas. Un horizonte superficial arenoso es ideal para la movilización de partículas.</p>		

text	var	color
Macroalgas	macroalgas	dark_green
https://fips.springeropen.com/articles/10.1186/s43094-020-00147-6		
Estacionalidad		
fuente:		
copernicus.		
posprocesar		
para ver		
también	secas	blue
temporada de		
secas		
VALERIA -		
Marisa		
proporciona el		
criterio		
Condición	condicion	green
Precipitación-		
fuente:		
copernicus.		
posprocesar		
para ver	lluvia_anual	blue
también		
temporada de		
secas		
VALERIA		
GastoEscurrimiento		
en		
cuencaAporte	gasto	blue
de sedimento		
terrígeno		
dirección de la		
ola (datos	dir_ola	blue
horarios 1940-		
2022)	https://confluence.ecmwf.int/display/CKB/ERA5%3A&#43;data&#43;documentation	

text	var	color
velocidad del viento (datos horarios 1940- 2022) https://confluence.ecmwf.int/display/CKB/ERA5%3A&#43;data&#43;documentationEventos		
de vientos extremos. Encontrar umbrales de viento y ver si favorecen la construcción o la destrucción de duna. ¿Velocidad de inicio de arrastre?WB_Atlas eólico mundial: https://globalwindatlas.info/es/PENDIENTE	vel_viento	blue
DEFINIR UMBRALES		
dirección del viento (datos horarios 1940- 2022) https://confluence.ecmwf.int/display/CKB/ERA5%3A&#43;data&#43;documentation	dir_viento	blue
Detección	inf_2	dark_green
Contexto	inf_1	blue
Otro	inf_4	violet
Intervención hunama	inf_3	orange
HumedadEn terminos de germinación y cobertura	humedad	dark_green

text	var	color
Checar pagina de idea del instituto de notas_2 gray geografíahttps://www.gits.igg.unam.mx/idea/descargaGaby Gómez		
Tasa de desplazamiento(Erosión- / acreción #43;) erosibilidad erosibilidad green	-	dark_green
Resolución raster100 m por lado (equivalete a 3 segundos)En versión "caso de resolucion cyan estudio"; serán pixeles de 10m por lado (equivale a 0.3 segundos).		
Ale:Presencia de especies indicadoras - dark_green sensibles a cambios en la costa:		
Aves marinas: chorlosAudubon - dark_green society y Conell		
Sitios de anidación de tortugas - dark_green marinas		

4 Taller

Taller de arranque del grupo de trabajo para unificar metodologías y determinar las variables a analizar

Lugar: [Edificio 17 del Instituto de Ingeniería CEMIE-Oceano](#), ciudad de México / en línea

Objetivo general:

Presentar la visión integral del proyecto, su origen, fundamentos teórico-metodológico y principios de ciencia abierta, para unificar criterios técnicos entre el grupo central de trabajo e identificar en conjunto las variables potenciales para los ejercicios de modelación.

Objetivos particulares:

1. Presentar el proyecto, sus etapas, los productos comprometidos y tiempos estimados.
2. Presentar las bases teórico-metodológicas del proyecto (redes bayesianas), así como los fundamentos de ciencia abierta (FAIR).
3. Identificar posibles variables a utilizar en el modelaje del proyecto, así como las fuentes de datos e información relevante en dos escalas (para el nivel nacional y un caso de estudio potencial a detalle en la Península de Yucatán).

Descripción de las actividades

Los talleres realizados el 29 y 31 de mayo contaron con la participación de 14 personas que conforman el grupo núcleo de trabajo del proyecto (ocho mujeres y seis hombres, tres participando en línea), provenientes del Instituto de Ingeniería de la UNAM, el Instituto de Ecología A. C. y la UAM – Xochimilco. Las listas de participantes se encuentran en este [Archivo](#).

El taller se realizó con un enfoque participativo, para lo cual, se preparó un [pizarrón digital](#) (plataforma Miro) y las presentaciones depositadas en el [blog de trabajo para el proyecto](#). El instrumento de planeación del taller (carta descriptiva) con la propuesta de actividades detalladas se encuentra en el anexo 2.

A continuación, se presentan la descripción del desarrollo de las actividades durante los tres días de trabajo.

Día 1.

1. Bienvenida y presentación

Al inicio del taller, el Dr. Octavio Pérez Maqueo dio una breve bienvenida al taller e informó que cada día se iba a pasar una lista de asistencia; además, se le pidió a cada persona señalar si brindaban su consentimiento libre e informado para el uso de imagen e información considerando la siguiente leyenda: “*Al marcar esta celda, usted da el consentimiento para que el proyecto ‘Estimación de la integridad ecosistémica de las costas arenosas mexicanas a través de técnicas de aprendizaje de máquina (CF-2023-G-1497)’ haga uso de materiales (fotos/videos) con su imagen, así como de la información derivada de los días de taller. Ninguna información privada o sensible será utilizada ni compartida fuera del marco del proyecto. Muchas gracias.*” (Ver listas en el anexo 1)

Posteriormente a la información de la lista y consentimiento, se invitó a cada participante a presentarse brevemente con su nombre e institución.

2. Presentación del contexto y origen del proyecto

El Dr. Pérez Maqueo desarrolló una presentación en torno a la pregunta ¿por qué creemos que es necesario un proyecto como este? Para ello, la exposición contó con los siguientes componentes:

Visión integral

- Caso particular de las costas
- Desarrollo de habilidades

¿Cómo sugerimos hacerlo?

- Co-construcción
- Ciencia Abierta y principios FAIR
- Análisis con base en propuesta del SEEA EA
- Flujos de trabajo (pipelines)

En esta primera presentación se destacaron algunos argumentos de la necesidad que dio origen al proyecto:

- Fortalecer iniciativas nacionales e internacionales que vemos como una alternativa viable hacia la sustentabilidad.
- Desarrollar metodologías que nos permitan a nosotros u otras personas rehacer mañana el análisis que hicimos hoy.
- Aprender sobre nuevas formas de análisis dada la capacidad de datos y cómputo actuales.
- Analizar a los socioecosistemas de manera integral incorporando las dependencias espaciales que caracteriza a estos sistemas.
- Implementar los resultados del proyecto en la toma de decisiones.

Las presentaciones están disponibles en el blog donde se subirán los materiales del proyecto:

<https://sw-costas-arenosas.netlify.app/presentaciones#category=presentaci%C3%B3n>

3. Descripción del proyecto y marco teórico

En la segunda parte de la mañana del día 1, se expusieron las generalidades del proyecto, el marco conceptual, los compromisos establecidos para las tres etapas del proyecto, presupuesto, avances y tiempos estimados.

El objetivo principal del proyecto es “integrar indicadores bióticos y abióticos en un índice de integridad ecosistémica, a través de un modelo causal bajo una aproximación de Inteligencia Artificial Interpretable (IAI)”:

- Nacional
- Incorporar la complejidad asociada a ecosistemas costeros
- Aprendizaje Automático Interpretable (redes bayesianas)

El resultado principal es proponer una metodología para analizar la integridad ecosistémica de las costas arenosas de México, con el fin de incorporarla a las cuentas nacionales de ecosistemas y promover su uso en distintos instrumentos de política pública.

La entrega de la primera etapa del proyecto es el día 27 de junio de 2024, por lo que dentro de los acuerdos se establecieron las tareas y responsables para cumplir plenamente con los compromisos (ver tabla de acuerdos, anexo 3).

Posteriormente, el Dr. Miguel Equihua brindó un panorama de los fundamentos teóricos en los que se sustenta el proyecto, principalmente explicando el proceso de análisis de interdependencias condicionadas a partir de las redes bayesianas (directed acyclic graph-DAG) (ver [blog](#)):

En la presentación se destacó la importancia de hacer la ciencia reproducible, lo que es un DAG, su anatomía, la terminología básica (nodo, arista, variable explicativa, respuesta, ancestros, descendientes), las reglas de independencia condicional y la separación condicional. Después del planteamiento teórico, se explicó cómo se busca construir el DAG de integridad costera. Finalmente se presentó el fundamento de la ciencia abierta y los principios FAIR.

De esta manera, las/os participantes reconocieron las bases del proyecto y su propuesta de desarrollo. Al terminar la exposición hubo un breve espacio para algunas dudas o inquietudes, incluyendo a las personas que atendieron en línea.

4. Presentación de la estrategia metodológica

La siguiente presentación buscó que el grupo cuente con la información base de la estrategia metodológica del taller, la cual sirvió también para realizar el ejercicio planeado para el día 2 de trabajo, en el cual se identificaron las variables y fuentes de información. La presentación contó con los siguientes elementos:

- De lo conceptual a lo analítico (de Miro a Nética)
- Enfoque causal y probabilístico
- Documentación y gestión de datos

A lo largo del día se plantearon diversas inquietudes, por ejemplo:

- Definir las fronteras del área de estudio.
- Definir si el grupo estaba en posibilidades de un estudio de caso a profundidad en la Península d Yucatán.
- Considerar si se considerará información de la cuenca a nivel superficial y/o a los acuíferos.
- Definir la temporalidad, qué años y si se toman las dos temporadas para las costas.
- Necesidad de llegar a las resoluciones espaciales para la información a nivel nacional y para un estudio de caso.

Además, se enlistó la información que podría ser relevante y algunas fuentes de datos que debían retomarse:

- Las imágenes que se usaron para el inventario de humedales, que tiene imágenes desde el 75 u 83 y del 2005-2006. Consultar a CONABIO.
- Las imágenes de los vuelos que tiene INEGI y/o SEMARNAT.
- Biblioteca de imágenes de ICA, donde incluso se cuenta con vuelos de los años sesenta (aerofotos).
- Acervo de la compañía mexicana de aerofoto, que se había depositado en el Instituto de Geografía-UNAM, sin embargo, como no estaban cuidados los rollos, ICA se los llevo. Los tiene la Fundación ICA.
- Las fotografías aéreas más antiguas que tomaron los EE. UU., las cuales están escaneadas. El equipo de Rodolfo Silva tiene acceso a esta información. Una parte ya está digitalizada, pero no toda está organizada, hay mucho trabajo pendiente.
- Información de la CFE.
- Las cartas geológico-mineras de todo el país. Los tiene el servicio geológico mexicano

El equipo facilitador tomó notas cada día de trabajo de estos puntos para llegar a definiciones en una tabla de acuerdos al final del taller (ver anexo 3).

Al final del día 1, se dio un espacio en plenaria de cierre de la jornada de trabajo y se propusieron los mecanismos de interacción para el día 2.

Día 2

5. Presentación del tema gestión de datos y presentación de variables

Al inicio del día 2, se continuó con el tema de gestión de datos y se presentaron las temáticas de variables existentes y nuevas variables, así como las sugerencias para documentar y gestionar datos en el marco del proyecto.

Posteriormente se brindó un ejemplo de variables y sus interacciones. De esta manera, el grupo conoció la metodología específica para definir variables e interacciones.

6. Ejercicio participativo de identificación de variables

a) Mecánica del ejercicio:

Una vez que se introdujo al grupo de trabajo a la metodología de identificación de variables, se realizó un ejercicio donde el grupo participó a través de notas digitales tipo post-it en un pizarrón digital en la plataforma Miro (ver Fig. 1). Para esto, el equipo facilitador había preparado previamente el pizarrón de trabajo, se propusieron variables que ya se habían identificado, además de presentar el ejemplo de un caso que se desarrolló en un proyecto previo.

4. Definir los parámetros técnicos para el intercambio de información entre el equipo de trabajo (resolución, formato, metadatos, etc.).

Resultados generados:

1. Mural digital con la lista de variables potenciales identificada por el grupo de trabajo, la cual incluye fuente de datos, la escala de información y la persona que propuso la variable y que podría gestionar el acceso a la información.
2. Lista de fuentes de información adicionales que puedan servir como insumo para el modelaje.
3. Tabla de acuerdos con tareas y responsables para la siguiente fase del proyecto.